ManipLLM: Embodied Multimodal Large Language Model for Object-Centric Robotic Manipulation

Year: 2023 | *Citations: 172* | *Authors: Xiaoqi Li, Mingxu Zhang, Yiran Geng*

Abstract

Robot manipulation relies on accurately predicting contact points and end-effector directions to ensure successful operation. However, learning-based robot manipulation, trained on a limited category within a simulator, often struggles to achieve generalizability, especially when confronted with extensive categories. Therefore, we introduce an innovative approach for robot manipulation that leverages the robust reasoning capabilities of Multimodal Large Language Models (MLLMs) to enhance the stability and generalization of manipulation. By fine-tuning the injected adapters, we preserve the inherent common sense and reasoning ability of the MLLMs while equipping them with the ability for manipulation. The fundamental insight lies in the introduced fine-tuning paradigm, encompassing object category understanding, affordance prior reasoning, and object-centric pose prediction to stimulate the reasoning ability of MLLM in manipulation. During inference, our approach utilizes an RGB image and text prompt to predict the end effector's pose in chain of thoughts. After the initial contact is established, an active impedance adaptation policy is introduced to plan the upcoming way-points in a closed-loop manner. Moreover, in real world, we design a test-time adaptation (TTA) strategy for manipulation to enable the model better adapt to the current real-world scene configuration. Experiments in simulator and real-world show the promising performance of Mani-pLLM. More details and demonstrations can be found at https://sites.google.com/view/manipllm.